Quotient Complexity of Closed Languages ⋆

Janusz Brzozowski[1], Galina Jirásková[2], and Chenglong Zou[1]

[1] David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1
{brzozo@,c2zou@student.math.}uwaterloo.ca
[2] Mathematical Institute, Slovak Academy of Sciences, Grešákova 6, 040 01 Košice, Slovakia
{jiraskov@saske.sk}

Abstract. A language L is prefix-closed if, whenever a word w is in L, then every prefix of w is also in L. We define suffix-, factor-, and subword-closed languages in an analogous way, where by subword we mean subsequence. We study the quotient complexity (usually called state complexity) of operations on prefix-, suffix-, factor-, and subword-closed languages. We find tight upper bounds on the complexity of the subword-closure of arbitrary languages, and on the complexity of boolean operations, concatenation, star, and reversal in each of the four classes of closed languages. We show that repeated application of positive closure and complement to a closed language results in at most four distinct languages, while Kleene closure and complement gives at most eight.

Keywords: automaton, closed, factor, language, prefix, quotient, regular operation, state complexity, subword, suffix, upper bound

1 Introduction

The *state complexity of a regular language* L is the number of states in the minimal deterministic finite automaton (dfa) recognizing L. The *state complexity of an operation* in a subclass $\mathcal{C}$ of regular languages is defined as the worst-case size of the minimal dfa accepting the language resulting from the operation, taken as a function of the quotient complexities of the operands in $\mathcal{C}$.

The first results for the state complexity of reversal are due to Mirkin [21] (1966), and of union, concatenation, and star, to Maslov [20] (1970). For a general discussion of state complexity see [5, 27] and the reference lists in those papers. In 1994 the complexity of concatenation, star, left and right quotients, reversal, intersection, and union in regular languages was examined in detail in [28]. The complexity of operations was also considered in several subclasses of regular languages: unary [23, 28], finite [27], cofinite [3], prefix-free [14], suffix-free [13], and ideal [7]. These studies show that the complexity can be significantly lower in a subclass than in the general case. Here we examine state complexity in the classes of prefix-, suffix-, factor-, and subword-closed regular languages.

⋆ This work was supported by the Natural Sciences and Engineering Research Council of Canada grant OGP0000871 and by VEGA grant 2/0111/09.

There are several reasons for considering closed languages. Subword-closed languages were studied in 1969 [12], and also in 1973 [25]. Suffix-closed languages were considered in 1974 [11], and later in [10, 15, 26]. Factor-closed languages, also called *factorial*, have received some attention, for example, in [2, 19]. Subword-closed languages were studied in [22]. The state complexities of the prefix-, suffix-, and factor-closure of a language were examined in [17]. Prefix-closed languages play a role in predictable semiautomata [8]. All four classes of closed languages were examined in [1], and decision problems for closed languages were studied in [9]. A language is a *left ideal* (respectively, *right, two-sided, all-sided ideal*) if $L = \Sigma^* L$, (respectively, $L = L\Sigma^*$, $L = \Sigma^* L \Sigma^*$, and $L = \Sigma^* \sqcup\!\sqcup L$, where $\Sigma^* \sqcup\!\sqcup L$ is the shuffle of Σ^* with L). Closed languages are related to ideal languages as follows [1]: A non-empty language is a right (left, two-sided, all-sided) ideal if and only its complement is a prefix (suffix, factor, subword)-closed language. Closed languages are defined by binary relations "is a prefix of" (respectively, "is a suffix of", "is a factor of", "is a subword of") [1], and are special cases of convex languages [1, 25]. The fact that the four classes of closed languages are related to each other permits us to obtain many complexity results using similar methods.

2 Quotient Complexity

If Σ is a non-empty finite *alphabet*, then Σ^* is the free monoid generated by Σ. A *word* is any element of Σ^*, and ε is the *empty word*. A *language* over Σ is any subset of Σ^*. The cardinality of a set S is denoted by $|S|$. If $w = uxv$ for some u, v, x in Σ^*, then u is a *prefix* of w, v is a *suffix* of w, and x is a *factor* of w. If $w = w_0 a_1 w_1 \cdots a_n w_n$, where $a_1, \ldots, a_n \in \Sigma$, and $w_0, \ldots, w_n \in \Sigma^*$, then the word $a_1 \cdots a_n$ is a *subword* of w.

A language L is *prefix-closed* if $w \in L$ implies that every prefix of w is also in L. In a similar way, we define *suffix-*, *factor-*, and *subword-closed* languages. A language is *closed* if it is prefix-, suffix-, factor-, or subword-closed.

The following set operations are defined on languages: *complement* ($\overline{L} = \Sigma^* \setminus L$), *union* ($K \cup L$), *intersection* ($K \cap L$), *difference* ($K \setminus L$), and *symmetric difference* ($K \oplus L$). A general *boolean operation* with two arguments is denoted by $K \circ L$. We also define the *product*, usually called *concatenation* or *catenation*, ($KL = \{w \in \Sigma^* \mid w = uv, u \in K, v \in L\}$), (Kleene) *star* ($L^* = \bigcup_{i \geq 0} L^i$), and *positive closure* ($L^+ = \bigcup_{i \geq 1} L^i$). The *reverse* w^R of a word w in Σ^* is defined as follows: $\varepsilon^R = \varepsilon$, and $(wa)^R = aw^R$. The *reverse* of a language L is denoted by L^R and is defined as $L^R = \{w^R \mid w \in L\}$.

Regular languages over an alphabet Σ are languages that can be obtained from the *set of basic languages* $\{\emptyset, \{\varepsilon\}\} \cup \{\{a\} \mid a \in \Sigma\}$, using a finite number of operations of union, product, and star. Such languages are usually denoted by regular expressions. If E is a regular expression, then $\mathcal{L}(E)$ is the language denoted by that expression. For example, $E = (\varepsilon \cup a)^* b$ denotes the language $\mathcal{L}(E) = (\{\varepsilon\} \cup \{a\})^* \{b\}$. We usually do not distinguish notationally between regular languages and regular expressions; the meaning is clear from the context.

A *deterministic finite automaton* (dfa) is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where Q is a set of *states*, Σ is the *alphabet*, $\delta : Q \times \Sigma \to Q$ is the *transition function*, q_0 is the *initial state*, and F is the set of *final* or *accepting states*. A *nondeterministic finite automaton* (nfa) is a quintuple $\mathcal{N} = (Q, \Sigma, \eta, Q_0, F)$, where Q, Σ, and F are as in a dfa, $\eta : Q \times \Sigma \to 2^Q$ is the *transition function* and $Q_0 \subseteq Q$ is the *set of initial states*. If η also allows ε, that is, $\eta : Q \times (\Sigma \cup \{\varepsilon\}) \to 2^Q$, we call $\mathcal{N}$ an *ε-nfa*.

Our approach to quotient complexity follows closely that of [5]. Since state complexity is a property of a language, it is more appropriately defined in language-theoretic terms. The *left quotient*, or simply *quotient,* of a language L by a word w is the language $L_w = \{x \in \Sigma^* \mid wx \in L\}$. The *quotient complexity* of L is the number of distinct quotients of L, and is denoted by $\kappa(L)$.

Quotients of regular languages [4, 5] can be computed as follows: First, the *ε-function* L^ε of a regular language L is $L^\varepsilon = \emptyset$ if $\varepsilon \notin L$ and $L^\varepsilon = \varepsilon$ if $\varepsilon \in L$. The quotient by a letter a in Σ is computed by induction: $b_a = \emptyset$ if $b \in \{\emptyset, \varepsilon\}$ or $b \in \Sigma$ and $b \neq a$, and $b_a = \varepsilon$ if $b = a$; $(\overline{L})_a = \overline{L_a}$; $(K \cup L)_a = K_a \cup L_a$; $(KL)_a = K_a L \cup K^\varepsilon L_a$; $(L^*)_a = L_a L^*$. The quotient by a word w in Σ^* is computed by induction on the length of w: $L_\varepsilon = L$ and $L_{wa} = (L_w)_a$. A quotient L_w is *accepting* if $\varepsilon \in L_w$; otherwise it is *rejecting*.

The *quotient automaton* of a regular language L is $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where $Q = \{L_w \mid w \in \Sigma^*\}$, $\delta(L_w, a) = L_{wa}$, $q_0 = L_\varepsilon = L$, and $F = \{L_w \mid (L_w)^\varepsilon = \varepsilon\}$. This is a minimal dfa for L; so quotient complexity of L equals the state complexity of L. However, there are some advantages to using quotients [5]. To simplify the notation, we write $(L_w)^\varepsilon$ as L_w^ε. Whenever convenient, we use the formulas given in the next proposition to establish upper bounds on quotient complexity.

Proposition 1 ([4, 5]). *If K and L are regular languages, then*

$$(\overline{L})_w = \overline{L_w}; \quad (K \circ L)_w = K_w \circ L_w. \tag{1}$$

$$(KL)_w = K_w L \cup K^\varepsilon L_w \cup \Big(\bigcup_{\substack{w=uv\\ u,v\in\Sigma^+}} K_u^\varepsilon L_v\Big). \tag{2}$$

$$(L^*)_\varepsilon = \varepsilon \cup LL^*, \quad (L^*)_w = \Big(L_w \cup \bigcup_{\substack{w=uv\\ u,v\in\Sigma^+}} (L^*)_u^\varepsilon L_v\Big) L^* \ \ \text{for } w \in \Sigma^+. \tag{3}$$

3 Closure Operations

Let $\trianglelefteq$ be a partial order on Σ^*; the *$\trianglelefteq$-closure* of a language L is the language ${}_{\trianglelefteq}L = \{x \in \Sigma^* \mid x \trianglelefteq w \text{ for some } w \in L\}$. We use $\leq$, $\preceq$, $\sqsubseteq$, $\Subset$ for the relations "is a prefix of", "is a suffix of", "is a factor of", "is a subword of", respectively.

The worst-case quotient complexity for closure was studied by Kao, Rampersad, and Shallit [17]. For suffix-closure, the bound $2^n - 1$ holds in case L does not have the empty quotient. We add the case where L has the empty quotient; here the bound is 2^{n-1}. Subword-closure was previously studied by Okhotin [22], but tight upper bounds were not established. Our next theorem solves this problem. For the sake of completeness, we provide all proofs.

Theorem 1 (Closure Operations). *Let $n \geq 2$. Let L be a regular language over an alphabet Σ with $\kappa(L) = n$. Let ${}_{\leq}L$, ${}_{\preceq}L$, ${}_{\sqsubseteq}L$, ${}_{\Subset}L$ be the prefix-, suffix-, factor-, and subword-closure of L, respectively. Then*

1. *$\kappa({}_{\leq}L) \leq n$, and the bound is tight if $|\Sigma| \geq 1$;*
2. *$\kappa({}_{\preceq}L) \leq 2^n - 1$ if L does not have empty quotient and $\kappa({}_{\preceq}L) \leq 2^{n-1}$ otherwise, and both bounds are tight if $|\Sigma| \geq 2$;*
3. *$\kappa({}_{\sqsubseteq}L) \leq 2^{n-1}$, and the bound is tight if $|\Sigma| \geq 2$;*
4. *$\kappa({}_{\Subset}L) \leq 2^{n-2} + 1$, and the bound is tight if $|\Sigma| \geq n - 2$.*

Proof. 1. Given a language L recognized by dfa $\mathcal{D}$, to get the dfa for its prefix-closure ${}_{\leq}L$, we need only make each non-empty state accepting. Thus $\kappa({}_{\leq}L) \leq n$. For tightness, consider the language $L = \{a^i \mid i \leq n - 2\}$. We have $\kappa({}_{\leq}L) = n$.

2. Having a quotient automaton of a language L, we can construct an nfa for its suffix-closure by making each non-empty state initial. The equivalent dfa has at most $2^n - 1$ states if L does not have the empty quotient (the empty set of states cannot be reached), and at most 2^{n-1} states otherwise.

To prove tightness, consider the language L defined by the quotient automaton shown in Fig. 1. Construct an nfa for the suffix-closure of L, by making all states initial. We show that the corresponding subset automaton has $2^n - 1$ reachable and pairwise inequivalent states. We prove reachability by induction

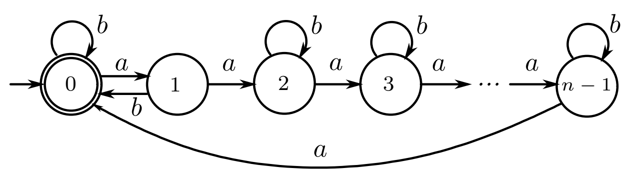

Fig. 1. Quotient automaton of a language L which does not have $\emptyset$.

on the size of subsets. The basis, $|S| = n$, holds since $\{0, 1, \ldots, n-1\}$ is the initial state. Assume that each set of size k is reachable, and let S be a set of size $k - 1$. If S contains 0 but does not contain 1, then it can be reached from the set $S \cup \{1\}$ of size k by b. If S contains both 0 and 1, then there is an i such that $i \in S$ and $i + 1 \notin S$. Then S can be reached from $\{(s - i) \bmod n \mid s \in S\}$ by a^i. The latter set contains 0 and does not contain 1, and so is reachable. If a non-empty S does not contain 0, then it can be reached from $\{s - \min S \mid s \in S\}$, which contains 0, by $a^{\min S}$. To prove inequivalence notice that the word a^{n-i} is accepted by the nfa only from state i for all $i = 0, 1, \ldots, n-1$. It turns out that all the states in the subset automaton are pairwise inequivalent.

Now consider the case where a language has the empty quotient. Let L be defined by the dfa of Fig. 2. Remove state $n - 1$ and all transitions going to it, and then construct an nfa as above. The proof of reachability of all non-empty subsets of $\{0, 1, \ldots, n-2\}$ is the same as above. The empty set is reached from $\{0\}$ by b. For inequivalence, the word $(ab)^n$ is accepted only from state 0, and the word $a^{n-1-i}(ab)^n$ is accepted only from state i for $i = 1, 2, \ldots, n-2$.

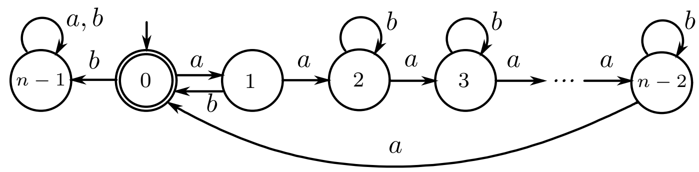

Fig. 2. Quotient automaton of a language L which has $\emptyset$.

3. Suppose we have the quotient automaton of a language L. To find an nfa for the factor closure $_\sqsubseteq L$, we make all non-empty states of the quotient automaton both accepting and initial, and delete the empty state. Hence the bound is 2^{n-1}. The language L defined by the quotient automaton of Fig. 2 meets the bound.

4. To get an ε-nfa for the subword-closure $_\Subset L$ from the quotient automaton of L, we remove the empty state (if there is no empty state, then $_\Subset L = \Sigma^*$), and add an ε-transition from state p to state q whenever there is a transition from p to q in the quotient automaton. Since the initial state can reach every non-empty state by ε-transitions, no other subset containing the initial state can be reached. Hence there are at most $2^{n-2}+1$ reachable subsets.

To prove tightness, if $n=2$, let $\Sigma=\{a,b\}$; then $L=a^*$ meets the bound. If $n \geq 3$, let $\Sigma=\{a_1,\dots,a_{n-2}\}$, and $L=\bigcup_{a_i\in\Sigma} a_i(\Sigma\setminus\{a_i\})^*$. Thus L consists of all words over Σ in which the first letter occurs exactly once. Let K be the subword-closure of L. Then $K = L \cup \{w \in \Sigma^* \mid \text{at least one letter is missing in } w\}$. For each boolean vector $b=(b_1,b_2,\dots,b_{n-2})$, define the word $w(b)=w_1w_2\cdots w_{n-2}$, in which $w_i=\varepsilon$ if $b_i=0$ and $w_i=a_i$ if $b_i=1$. Consider ε, and each word $a_1w(b)$. All the quotients of K by these $2^{n-2}+1$ words are distinct: For each binary vector b, we have $a_1a_2\cdots a_{n-2} \in K_\varepsilon \setminus K_{a_1w(b)}$. Let b and b' be two different vectors with $b_i=0$ and $b'_i=1$. Then we have $a_1a_2\cdots a_{i-1}a_{i+1}a_{i+2}\cdots a_{n-2} \in K_{a_1w(b)} \setminus K_{a_1w(b')}$. Thus all quotients are distinct, and so $\kappa(K) \geq 2^{n-2}+1$. □

4 Basic Operations on Closed Languages

Now we study the quotient complexity of operations on closed languages. For regular languages, the following bounds are known [20, 21, 28]: mn for boolean operations, $m2^n - 2^{n-1}$ for product, $2^{n-1}+2^{n-2}$ for star, and 2^n for reversal. The bounds for closed languages are smaller in most cases. The bounds for boolean operations and reversal follow from the results on ideal languages [7].

Theorem 2 (Boolean Operations). *Let K and L be prefix-closed (or factor-closed, or subword-closed) languages with $\kappa(K)=m$ and $\kappa(L)=n$. Then*
1. $\kappa(K\cap L) \leq mn-(m+n-2)$,
2. $\kappa(K\cup L), \kappa(K\oplus L) \leq mn$,
3. $\kappa(K\setminus L) \leq mn-(n-1)$.
For suffix-closed languages, $\kappa(K\circ L) \leq mn$. All bounds are tight if $|\Sigma| \geq 2$ except for the union and difference of suffix-closed languages where $|\Sigma| \geq 4$ is required.

Proof. The complement of a prefix-closed (suffix-, factor-, or subword-closed) language is a right (respectively, left, two-sided, all-sided) ideal. We get all the results using De Morgan's laws and the results from [7]. □

Remark 1. If L is prefix-closed, then either $L = \Sigma^*$ or L has the empty quotient. Moreover, each quotient of L is either accepting or empty.

Remark 2. For a suffix-closed language L, if v is a suffix of w then $L_w \subseteq L_v$. In particular, $L_w \subseteq L_\varepsilon = L$ for each word w in Σ^*.

Theorem 3 (Product). *Let $m, n \geq 2$. Let K and L be closed languages with $\kappa(K) = m$, $\kappa(L) = n$, and let k be the number of accepting quotients of K.*
1. If K and L are prefix-closed, then $\kappa(KL) \leq (m+1) \cdot 2^{n-2}$.
2. If K and L are suffix-closed, then $\kappa(KL) \leq (m-k)n + k$.
3. If K and L are both factor- or both subword-closed, then $\kappa(KL) \leq m + n - 1$.
The first two bounds are tight if $|\Sigma| \geq 3$, and the third bound is tight if $|\Sigma| \geq 2$. If $\kappa(K) = 1$ or $\kappa(L) = 1$, then $\kappa(KL) = 1$.

Proof. If $m = 1$, then $K = \emptyset$ or $K = \Sigma^*$; so $KL = \emptyset$ or $KL = \Sigma^*$, for if $L \neq \emptyset$, then $\varepsilon \in L$. Thus $\kappa(KL) = 1$. The case $n = 1$ is similar. Now let $m, n \geq 2$.

1. If K and L are prefix-closed, then $\varepsilon \in K$ and by Remark 1 both languages have the empty quotient. The quotient $(KL)_w$ is given by Equation (2). If K_w is accepting, then L is always in the union, and there are 2^{n-2} non-empty subsets of non-empty quotients of L that can be added. Since there are $m-1$ accepting quotients of K, there are $(m-1)2^{n-2}$ such quotients of KL. If K_w is rejecting, then 2^{n-1} subsets of non-empty quotients of L can be added.

For tightness, consider prefix-closed languages K and L defined by the quotient automata of Fig. 3 (if $n = 2$, then $L = \{a, c\}^*$). Construct an ε-nfa for

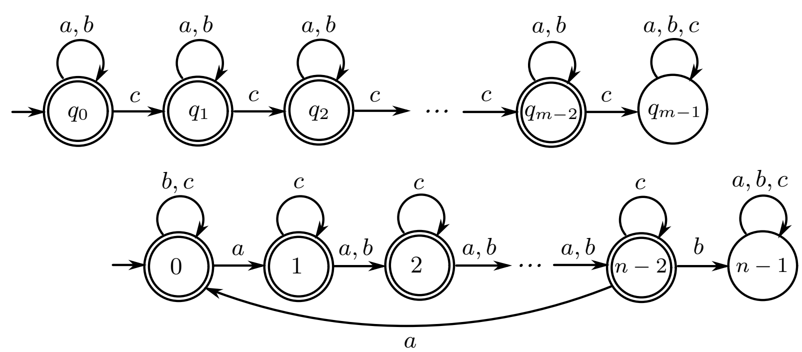

Fig. 3. Quotient automata of prefix-closed languages K and L.

the language KL from these quotient automata by adding an ε-transition from states $q_0, q_1, \ldots, q_{m-2}$ to state 0. The initial state of the nfa is q_0, and the accepting states are $0, 1, \ldots, n-2$. We show that there are $(m+1) \cdot 2^{n-2}$ reachable and pairwise inequivalent states in the corresponding subset automaton.

State $\{q_0, 0\}$ is the initial state, and each state $\{q_0, 0, i_1, i_2, \dots, i_k\}$, where $1 \le i_1 < i_2 < \dots < i_k \le n-2$, is reached from state $\{q_0, 0, i_2 - i_1, \dots, i_k - i_1\}$ by $ab^{i_1 - 1}$. For each subset S of $\{0, 1, \dots, n-2\}$ containing 0, each state $\{q_i\} \cup S$ with $1 \le i \le m-1$ is reached from $\{q_0\} \cup S$ by c^i. If a non-empty S does not contain 0, then $\{q_{m-1}\} \cup S$ is reached from $\{q_{m-1}\} \cup \{s - \min S \mid s \in S\}$, which contains 0, by $a^{\min S}$. State $\{q_{m-1}, n-1\}$ is reached from $\{q_{m-1}, n-2\}$ by b.

To prove inequivalence, notice that the word b^n is accepted by the quotient automaton for L only from state 0, and the word $a^{n-1-i}b^n$ only from state i ($1 \le i \le n-2$). It turns out that two different states $\{q_{m-1}\} \cup S$ and $\{q_{m-1}\} \cup T$ are inequivalent. It follows that states $\{q_i\} \cup S$ and $\{q_i\} \cup T$ are inequivalent as well. States $\{q_i\} \cup S$ and $\{q_j\} \cup T$ with $i < j$ can be distinguished by $c^{m-1-j}b^n ab^n$.

2. If K and L are suffix-closed, then, by Remark 2, for each word w in Σ^* and u, v in Σ^+, we have $(KL)_w = K_w L \cup K^\varepsilon L_w \cup (\bigcup_{w=uv} K_u^\varepsilon L_v) = K_w L \cup L_x$ for some suffix x of w. If K_w is a rejecting quotient, there are at most $(m-k)n$ such quotients. If K_w is accepting, then $\varepsilon \in K_w$, and since $L_x \subseteq L_\varepsilon = L \subseteq K_w L$, we have $(KL)_w = K_w L$. There are at most k such quotients. Therefore there are at most $(m-k)n + k$ quotients in total.

To prove tightness, let K and L be ternary suffix-closed languages defined by the quotient automata of Fig. 4. Consider the words $\varepsilon = a^0 b^0$, and $a^i b^j$ with

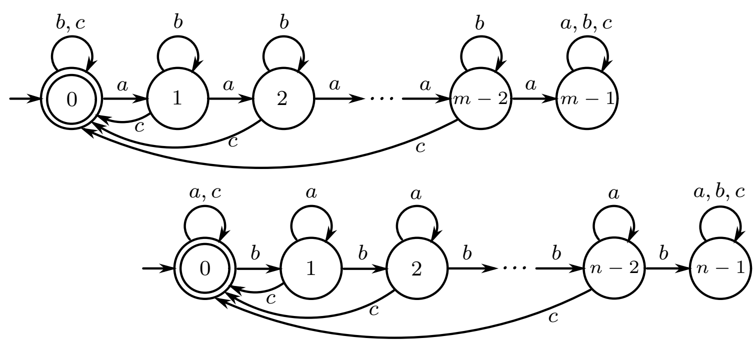

Fig. 4. Quotient automata of suffix-closed languages K and L.

$1 \le i \le m-1$ and $0 \le j \le n-1$. Let us show that all the quotients of KL by these words are distinct. Let $(i, j) \ne (k, \ell)$, and let $x = a^i b^j$ and $y = a^k b^\ell$. If $i < k$, take $z = a^{m-1-k} b^n c$. Then xz is in KL, while yz is not, and so $z \in (KL)_x \setminus (KL)_y$. If $i = k$ and $j < \ell$, take $z = a^m b^{n-1-\ell} c$. We again have $z \in (KL)_x \setminus (KL)_y$.

Notice that, if the quotients K_{a^i} with $0 \le i \le k-1$ are accepting, then the resulting product has quotient complexity $(m-k)n + k$.

3. It suffices to derive the bound for factor-closed languages, since every subword-closed language is also factor-closed. Since factor-closed languages are suffix-closed, $\kappa(KL) \le (m-k)n + k$. The language K has at most one rejecting quotient, because it is prefix-closed. Thus, $k = m-1$ and $\kappa(KL) \le m + n - 1$.

For tightness, let $K = \{w \in \{a,b\}^* \mid a^{m-1}$ is not a subword of $w\}$ and $L = \{w \in \{a,b\}^* \mid b^{n-1}$ is not a subword of $w\}$. The languages are subword-closed and $\kappa(K) = m$ and $\kappa(L) = n$. Consider the word $w = a^{m-1}b^{n-1}$. This word is not in the product KL. However, removing any non-empty subword from w results in a word in KL. Therefore, $\kappa(KL) \geq m+n-1$. □

Theorem 4 (Star). *Let $n \geq 2$. Let L be a closed language with $\kappa(L) = n$.*
1. If L is prefix-closed, then $\kappa(L^*) \leq 2^{n-2}+1$.
2. If L is suffix-closed, then $\kappa(L^*) = n$ if $L = L^*$, and $\kappa(L^*) \leq n-1$ if $L \neq L^*$.
3. If L is factor- or subword-closed, then $\kappa(L^*) \leq 2$.
The first bound is tight if $|\Sigma| \geq 3$, and all the other bounds are tight if $|\Sigma| \geq 2$. If $\kappa(L) = 1$, then $\kappa(L^*) \leq 2$.

Proof. 1. For every non-empty word w, the quotient $(L^*)_w$ is given by Equation (3). If L is prefix-closed, then so is L^* and $(L^*)_w$. Thus, if $(L^*)_w$ is non-empty, then it contains ε. Hence $(L^*)_w \supseteq L^* \supseteq L$. Since $\emptyset$ and L are always contained in every non-empty quotient of L^*, there are at most 2^{n-2} non-empty quotients of L^*. Since there is at most one empty quotient, there are at most $2^{n-2}+1$ quotients in total. The quotient $(L^*)_\varepsilon$ has already been counted, since L is closed and $\varepsilon \in L$ implies $(L^*)_\varepsilon = LL^*$, which has the form of Equation (3).

If $n = 1$ and $n = 2$, the bound 2 is met by $L = \emptyset$ and $L = \varepsilon$, respectively. Now let $n \geq 3$ and let L be the prefix-closed language defined by the dfa of Fig. 5; transitions not depicted in the figure go to state $n-1$. Construct an ε-nfa for

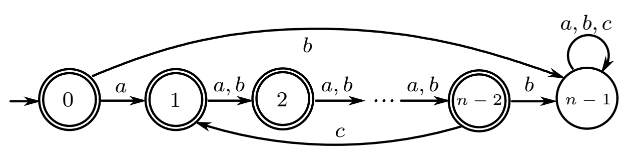

Fig. 5. Quotient automaton of prefix-closed language L.

L^* by removing state $n-1$ and adding an ε-transition from all the remaining states to the initial state. Let us show that $2^{n-2}+1$ states are reachable and pairwise inequivalent in the corresponding subset automaton.

We first prove that each subset of $\{0,1,\ldots,n-2\}$ containing state 0 is reachable. The proof is by induction on the size of the subsets. The basis, $|S| = 1$, holds since $\{0\}$ is the initial state of the subset automaton. Assume that each set of size k containing 0 is reachable, and let $S = \{0, i_1, i_2, \ldots, i_k\}$, where $0 < i_1 < i_2 < \cdots < i_k \leq n-2$, be a set of size $k+1$. Then S is reached from the set $\{0, i_2 - i_1, \ldots, i_k - i_1\}$ of size k by ab^{i_1-1}. Since the latter set is reachable by the induction hypothesis, the set S is reachable as well. The empty set can be reached from $\{0\}$ by b, and we have $2^{n-2}+1$ reachable states. To prove inequivalence notice that b^{n-3} is accepted by the nfa only from state 1, and each word $b^{n-2-i}cb^{n-3}$ $(2 \leq i \leq n-2)$, only from state i.

2. For a non-empty suffix-closed language L, the quotient $(L^*)_\varepsilon$ is LL^*, which is of the same form as the quotients by a non-empty word w in Equation (3), $(L^*)_w = (L_w \cup L_{v_1} \cup \cdots \cup L_{v_k})L^*$, where the v_i are suffixes of w, and v_k is the shortest. By Remark 2, $(L^*)_w = L_{v_k}L^*$. There are at most n such quotients. If $L \neq L^*$ for a non-empty suffix-closed language L, then there must be two words x, y in L such that $xy \notin L$. Hence $y \in L_\varepsilon \setminus L_x$, and so $L_\varepsilon \neq L_x$. However, since $\varepsilon \in L_x$ and L^* is suffix-closed, we have $(L^*)_\varepsilon = L^* \subseteq L_xL^* \subseteq (L^*)_x \subseteq (L^*)_\varepsilon$, and so $(L^*)_\varepsilon = (L^*)_x$. It turns out that $\kappa(L^*) \leq n-1$.

For $n = 1$, $L = \emptyset$ and for $n = 2$, $L = \varepsilon$ meet the bound 2. Let $n \geq 3$. If $L = (a \cup ba^{n-2})^*$, then L is suffix-closed, $\kappa(L) = n$, and $L^* = L$. If $L = \varepsilon \cup \bigcup_{i=0}^{n-3} a^i b$, then L is suffix-closed, $\kappa(L) = n$, $L^* = (\bigcup_{i=0}^{n-3} a^i b)^*$, and $\kappa(L^*) = n-1$.

3. If each letter in Σ appears in some word of a factor-closed language L, then $L^* = \Sigma^*$ and $\kappa(L^*) = 1$. Otherwise, $\kappa(L^*) = 2$. The bound is met by subword-closed language $L = \{w \in \{a,b\}^* \mid w = a^i \text{ and } 0 \leq i \leq n-2\}$. □

Since the operation of reversal commutes with complementation, the next theorem follows from the results on ideal languages [7].

Theorem 5 (Reversal). *Let $n \geq 2$. Let L be a closed language with $\kappa(L) = n$.*
1. If L is prefix-closed, then $\kappa(L^R) \leq 2^{n-1}$. The bound is tight if $|\Sigma| \geq 2$.
2. If L is suffix-closed, then $\kappa(L^R) \leq 2^{n-1} + 1$. The bound is tight if $|\Sigma| \geq 3$.
3. If L is factor-closed, then $\kappa(L^R) \leq 2^{n-2} + 1$. The bound is tight if $|\Sigma| \geq 3$.
4. If L is subword-closed, then $\kappa(L^R) \leq 2^{n-2}+1$. The bound is tight if $|\Sigma| \geq 2n$.
If $\kappa(L) = 1$, then $\kappa(L^R) = 1$. □

Unary Languages: Unary languages have special properties because the product of unary languages is commutative. The classes of prefix-closed, suffix-closed, factor-closed, and subword-closed unary languages all coincide. If a unary closed language L is finite, then either it is empty and so $\kappa(L) = 1$, or has the form $\{a^i \mid i \leq n-2\}$ and then $\kappa(L) = n$. If L is infinite, then $L = a^*$ and $\kappa(L) = 1$. The bounds for unary languages are in Tables 1 and 2 on page 11.

5 Kuratowski Algebras Generated by Closed Regular Languages

A theorem of Kuratowski [18] states that, given a topological space, at most 14 distinct sets can be produced by repeatedly applying the operations of closure and complement to a given set. A closure operation on a set S is an operation $\square : 2^S \to 2^S$ satisfying the following conditions for any subsets X, Y of S: (1) $X \subseteq X^\square$, (2) $X \subseteq Y$ implies $X^\square \subseteq Y^\square$, (3) $X^{\square\square} \subseteq X^\square$.

Kuratowski's theorem was studied in the setting of formal languages in [6]. Positive closure and Kleene closure (star) are both closure operations. It then follows that at most 10 distinct languages can be produced by repeatedly applying the operations of positive closure and complement to a given language, and at most 14 distinct languages can be produced with Kleene closure instead of positive closure. We consider here the case where the given language is closed and regular, and give upper bounds on the quotient complexity of the resulting languages. We denote the complement of a language L by L^-, the positive closure of the complement of L by L^{-+}, etc.

We begin with positive closure. Let L be a $\trianglelefteq$-closed language not equal to Σ^*. Then L^- is an ideal, and $L^{-+} = L^-$. In addition, L^+ is also $\trianglelefteq$-closed, so $L^{+-+} = L^{+-}$. Hence there are at most 4 distinct languages that can be produced with positive closure and complementation.

Theorem 6. *The worst-case complexities in the 4-element algebra generated by a closed language L with $\kappa(L) = n$ under positive closure and complement are as follows: $\kappa(L) = \kappa(L^-) = n$, $\kappa(L^+) = \kappa(L^{+-}) = f(n)$, where $f(n)$ is $2^{n-2}+1$ for prefix-closed languages, $n-1$ for suffix-closed languages, and 2 for factor- and subword-closed languages. There exist closed languages that meet these bounds.*

Proof. Since $L^+ = L^*$ for a non-empty closed language we have $\kappa(L^+) = \kappa(L^*)$, and the upper bounds $f(n)$ follow from our results on the quotient complexity of the star operation; in the case of suffix-closed languages, to get a 4-element algebra we need $L \neq L^*$. All the languages that we have used in Theorem 4 to prove tightness can be used as examples meeting the bound $f(n)$. $\square$

The case of Kleene closure is similar. Let L be a non-empty $\trianglelefteq$-closed language that is not equal to Σ^*. Then the language L^- is an ideal and L^- does not contain ε. Thus $L^{-*} = L^- \cup \varepsilon$ and $L^{-*-} = L \setminus \varepsilon$, which gives at most four languages thus far. Now $L^* = (L \setminus \varepsilon)^*$, and the language L^* is also $\trianglelefteq$-closed. By the previous reasoning, we have at most four additional languages, giving a total of eight languages as the upper bound. The 8-element algebras are of the form $(L, L^-, L^{-*} = L^- \cup \varepsilon, L^{-*-} = L \setminus \varepsilon, L^*, L^{*-}, L^{*-*} = L^{*-} \cup \varepsilon, L^{*-*-} = L^* \setminus \varepsilon)$.

Theorem 7. *The worst-case quotient complexities in the 8-element algebra generated by a closed language L with $\kappa(L) = n$ under star and complement are as follows: $\kappa(L) = \kappa(L^-) = n$, $\kappa(L^*) = \kappa(L^{*-}) = f(n)$, $\kappa(L^{*-*}) = \kappa(L^{*-*-}) = f(n) + 1$, $\kappa(L^{-*}) = \kappa(L^{-*-}) = n + 1$, where $f(n)$ is $2^{n-2} + 1$ for prefix-closed languages, $n-1$ for suffix-closed languages, and 2 for factor-and subword-closed languages. Moreover, there exist closed languages that meet these bounds.*

Proof. Since $L^{-*-} = L \setminus \varepsilon$ and $L^{*-*-} = L^* \setminus \varepsilon$ we have $\kappa(L^{-*-}) \leq n + 1$ and $\kappa(L^{*-*-}) \leq f(n) + 1$. In the case of suffix-closed languages, since L must be distinct from L^*, we have $f(n) = n - 1$ by Theorem 4.

1. Let L be the prefix-closed language defined by the quotient automaton in Fig. 5 on page 8; then L meets the upper bound on star. Add a loop with a new letter d in each state and denote the resulting language by K. Then K is a prefix-closed language with $\kappa(K) = n$ and $\kappa(K \setminus \varepsilon) = n + 1$. Next we have $\kappa(K^*) = \kappa(L^*) = 2^{n-2} + 1$ and $\kappa(K^* \setminus \varepsilon) = 2^{n-2} + 2$.

2. Let $L = b^* \cup \bigcup_{i=1}^{n-3} b^* a^i b$. Then L is a suffix-closed language with $\kappa(L) = n$ and $\kappa(L \setminus \varepsilon) = n + 1$. Next $\kappa(L^*) = n - 1$ and $\kappa(L^* \setminus \varepsilon) = n$.

3. Let $L = \{w \in \{a, b, c\}^* \mid w = b^* a^i \text{ and } 0 \leq i \leq n - 2\}$. Then L is a subword-closed language with $\kappa(L) = n$ and $\kappa(L \setminus \varepsilon) = n+1$. Next $L^* = \{a, b\}^*$, and so $\kappa(L^*) = 2$ and $\kappa(L^* \setminus \varepsilon) = 3$. $\square$

6 Conclusions

Tables 1 and 2 summarize our complexity results. The complexities for regular languages are from [16, 20, 21, 28], except those for difference and symmetric difference, which are from [5]. The bounds for boolean operations and reversal of closed languages are direct consequences of the results in [7]. In Table 2, k is the number of accepting quotients of K; the results for prefix-, suffix-, and factor-closure are from [17]. The tables also show the size of the alphabet of the witness languages. In all cases when the size of the alphabet is more than two, we do not know whether the bounds are tight for a smaller alphabet.

	$K \cup L$	$\|\Sigma\|$	$K \cap L$	$\|\Sigma\|$	$K \setminus L$	$\|\Sigma\|$	$K \oplus L$	$\|\Sigma\|$
unary closed	$\max(m,n)$	1	$\min(m,n)$	1	m	1	$\max(m,n)$	1
prefix-, factor-, subword-closed	mn	2	$mn-m-n+2$	2	$mn-n+1$	2	mn	2
suffix-closed	mn	4	mn	2	mn	4	mn	2
regular	mn	2	mn	2	mn	2	mn	2

Table 1. Bounds on quotient complexity of boolean operations.

	$\trianglelefteq L$	$\|\Sigma\|$	KL	$\|\Sigma\|$	K^*	$\|\Sigma\|$	K^R	$\|\Sigma\|$
unary closed	n	1	$m+n-2$	1	2	1	n	1
prefix–closed	n	1	$(m+1)2^{n-2}$	3	$2^{n-2}+1$	3	2^{n-1}	2
suffix-closed	2^n-1	2	$(m-k)n+k$	3	n	2	$2^{n-1}+1$	3
factor-closed	2^{n-1}	2	$m+n-1$	2	2	2	$2^{n-2}+1$	3
subword-closed	$2^{n-2}+1$	$n-2$	$m+n-1$	2	2	2	$2^{n-2}+1$	$2n$
regular	–	–	$m2^n-k2^{n-1}$	2	$2^{n-1}+2^{n-k-1}$	2	2^n	2

Table 2. Bounds on quotient complexity of closure, product, star and reversal.

References

1. Ang, T., Brzozowski, J.: Languages convex with respect to binary relations, and their closure properties. Acta Cybernet. 19, 445–464 (2009)
2. Avgustinovich, S.V., Frid, A.E.: A unique decomposition theorem for factorial languages. Internat. J. Algebra Comput. 15, 149–160 (2005)
3. Bassino, F., Giambruno, L., Nicaud, C.: Complexity of operations on cofinite languages. In: López-Ortiz, A. (ed.) LATIN 2010. LNCS, vol. 6034, pp. 222–233. Springer-Verlag Berlin Heidelberg (2010)
4. Brzozowski, J.: Derivatives of regular expressions. J. ACM 11, 481–494 (1964)
5. Brzozowski, J.: Quotient complexity of regular languages. In: Dassow, J., Pighizzini, G., Truthe, B. (eds.) 11th International Workshop on Descriptional Complexity of Formal Systems, DCFS 2009, pp. 25–42. Otto-von-Guericke-Universität, Magdeburg, Germany (2009) http://arxiv.org/abs/0907.4547

6. Brzozowski, J., Grant, E., Shallit, J.: Closures in formal languages and Kuratowski's theorem. In: Diekert, V., Nowotka, D. (eds.) DLT 2009. LNCS, vol. 5583, pp. 125–144. Springer-Verlag Berlin Heidelberg (2009)
7. Brzozowski, J., Jirásková, G., Li, B.: Quotient complexity of ideal languages. In: López-Ortiz, A. (ed.) LATIN 2010. LNCS, vol. 6034, pp. 208–211. Springer-Verlag Berlin Heidelberg (2010). Full paper at `http://arxiv.org/abs/0908.2083`
8. Brzozowski, J., Santean, N.: Predictable semiautomata. Theoret. Comput. Sci. 410, 3236–3249 (2009)
9. Brzozowski, J., Shallit, J., Xu, Z.: Decision procedures for convex languages. In: Dediu, A., Ionescu, A., Martin-Vide, C. (eds.) LATA 2009. LNCS, vol. 5457, pp. 247–258. Springer-Verlag Berlin Heidelberg (2009)
10. Galil, Z., Simon, J.: A note on multiple-entry finite automata. J. Comput. System Sci. 12, 350–351 (1976)
11. Gill, A., Kou, L.T.: Multiple-entry finite automata. J. Comput. System Sci. 9, 1–19 (1974)
12. Haines, L.H.: On free monoids partially ordered by embedding. J. Combin. Theory 6, 94–98 (1969)
13. Han, Yo-S., Salomaa, K.: State complexity of basic operations on suffix-free regular languages. Theoret. Comput. Sci. 410, 2537–2548 (2009)
14. Han, Yo-S., Salomaa, K., Wood, D.: Operational state complexity of prefix-free regular languages. In: Automata, Formal Languages, and Related Topics, pp. 99–115. University of Szeged, Hungary (2009)
15. Holzer, M., Salomaa, K., Yu, S.: On the state complexity of k-entry deterministic finite automata. J. Autom. Lang. Comb. 6, 453–466 (2001)
16. Jirásek, J., Jirásková, G., Szabari, A.: State complexity of concatenation and complementation. Internat. J. Found. Comput. Sci. 16, 511–529 (2005)
17. Kao, J-Y., Rampersad, N., Shallit, J.: On NFAs where all states are final, initial, or both. Theoret. Comput. Sci. 410, 5010–5021 (2009)
18. Kuratowski, C.: Sur l'opération $\overline{A}$ de l'analysis situs. Fund. Math. 3, 182–199 (1922)
19. de Luca, A., Varricchio, S.: Some combinatorial properties of factorial languages. In: Capocelli, R. (ed.) Sequences, pp. 258–266. Springer (1990)
20. Maslov, A.N.: Estimates of the number of states of finite automata. Dokl. Akad. Nauk SSSR 194, 1266–1268 (1970) (in Russian). English translation: Soviet Math. Dokl. 11, 1373–1375 (1970)
21. Mirkin, B.G.: On dual automata. Kibernetika (Kiev) 2, 7–10 (1966) (in Russian). English translation: Cybernetics 2, 6–9 (1966)
22. Okhotin. A: On the state complexity of scattered substrings and superstrings. Turku Centre for Computer Science Technical Report No. 849 (2007)
23. Pighizzini, G., Shallit, J.: Unary language operations, state complexity and Jacobsthal's function. Internat. J. Found. Comput. Sci. 13, 145–159 (2002)
24. Salomaa, A., Wood, D., Yu, S.: On the state complexity of reversals of regular languages. Theoret. Comput. Sci. 320, 315–329 (2004)
25. Thierrin, G.: Convex languages. In: Nivat, M. (ed.) Automata, Languages and Programming, pp. 481–492. North-Holland (1973)
26. Veloso, P.A.S., Gill, A.: Some remarks on multiple-entry finite automata. J. Comput. System Sci. 18, 304–306 (1979)
27. Yu., S.: State complexity of regular languages. J. Autom. Lang. Comb. 6, 221–234 (2001)
28. Yu, S., Zhuang, Q., Salomaa, K.: The state complexities of some basic operations on regular languages. Theoret. Comput. Sci. 125, 315–328 (1994)